Mining the digital universe of data to develop personalized cancer therapies

August 12, 2013

Icahn School of Medicine at Mount Sinai
Disclosures

• I am on the Scientific Advisory Board for
  – Pacific Biosciences
  – Numedii
  – StationX
  – Spiral Genetics
  – Berg Pharmaceuticals
  – Ingenuity
  – GNS Healthcare

• I am on the Board of Directors for
  – Sage Bionetworks
  – While Biome
Disclosures

• Given the apparent rampant use of performance enhancing drugs in sports:
Poll results: look who’s doping

In January, Nature launched an informal survey into readers’ use of cognition-enhancing drugs. Brendan Maher has waded through the results and found large-scale use and a mix of attitudes towards the drugs.

- I used no performance enhancing drugs to carry out any of the research I will discuss today

20% are abusing
Considering the digital universe of data to better diagnose and treat patients.
We need to be able to leverage the digital universe of information to best solve the most challenging problems.

1.8 ZETTABYTES

(1.8 trillion gigabytes of information will be created and replicated in 2011; growth continues to accelerate – factor of 9 growth in last 5 years)
Being masters of really big data now critical for biomedical research (TB → PB → EB → ZB)

Cost per Raw Megabase of DNA Sequence

Organisms | Tissues | Single cells
---|---|---

Single cell, real-time, continuous?
Big Data Warehouses at Medical Centers like Mount Sinai Contain Virtually All Facts And Transaction Records For Millions of Patients

Institute for Personalized Medicine at Mount Sinai
Multiscale measures of patients now available through efforts like Mount Sinai’s Biobank (>25,000 *identified* patients and growing fast)
These technologies are enabling scoring of very large-scale, high-dimensional data on individuals for low cost.

- Modified and unmodified DNA
- Modified and unmodified coding and non-coding RNA
- Phosphorylated and unphosphorylated proteins
- Metabolites
That promise to enable the construction of molecular networks that define the biological processes that comprise living systems.
Integrating data to build predictive models of living systems

( - DNA,  - RNA,  - Protein,  - Metabolite)
An integrative genomics approach to the reconstruction of gene networks in segregating populations

J. Zhu, a P.Y. Lum, a J. Lamb, a D. GuhaThakurta, a S.W. Edwards, a R. Thieringer, b J.P. Berger, c M.S. Wu, d J. Thompson, e A.B. Sachs, a and E.E. Schadt a
Leveraging DNA variations as a perturbation source is key to inferring causality.
Understanding the network architecture critical for understanding how information flows through it
Stratifying patient populations
Integrating data to build predictive models of complex disease and drug response phenotypes.

Input

DNA regulatory proteins

RNA networks

Protein network

Molecular signaling

Integrative network model of genes and traits
Organizing 163 genetic loci for IBD

Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease

Problem: How do you make sense of 163 loci to understand a complex disease like IBD?
Connections between diseases and tissues: IBD network driving Alzheimer’s

Building networks from 500 prefrontal cortex samples
Constructing the co-expression networks

“Normal” versus LOAD Networks

<table>
<thead>
<tr>
<th>Pink</th>
<th>LOAD</th>
<th>Green yellow</th>
<th>Light green</th>
<th>Chartreuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>GST activity</td>
<td>(92.3, &lt;0.001)</td>
<td>Unfolded protein</td>
<td>Nerve myelination</td>
<td>Glucosyl transferase</td>
</tr>
<tr>
<td>Khaki</td>
<td>GABA biosynthesis</td>
<td>(0.3, &lt;0.001)</td>
<td></td>
<td>(2.3, 0.02)</td>
</tr>
<tr>
<td>Orchid</td>
<td>Phototransduction</td>
<td>(77.6, &lt;0.001)</td>
<td>Dynein complex</td>
<td>Ribosome</td>
</tr>
<tr>
<td>Chocolate</td>
<td>(2.1, &lt;0.002)</td>
<td></td>
<td></td>
<td>(24.9, &lt;0.001)</td>
</tr>
<tr>
<td>Red 3</td>
<td>Ribosome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khaki</td>
<td>GABA biosynthesis</td>
<td>(0.3, 0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green 4</td>
<td>Cell cycle</td>
<td>(0.3, &lt;0.001)</td>
<td>Neurogenesis</td>
<td></td>
</tr>
<tr>
<td>Midnight blue</td>
<td></td>
<td></td>
<td>(8.0, &lt;0.001)</td>
<td></td>
</tr>
<tr>
<td>Yellow</td>
<td>Immune functions</td>
<td></td>
<td></td>
<td>(2.0, &lt;0.001)</td>
</tr>
<tr>
<td>Cyan 2</td>
<td>Muscle contraction</td>
<td>(0.9, 0.23)</td>
<td>Olfactory perception</td>
<td></td>
</tr>
<tr>
<td>Brown 2</td>
<td></td>
<td></td>
<td>(25.5, &lt;0.001)</td>
<td></td>
</tr>
<tr>
<td>Firebrick</td>
<td>Neutrophil signaling</td>
<td></td>
<td></td>
<td>(0.3, &lt;0.001)</td>
</tr>
<tr>
<td>Tan</td>
<td>Extracellular matrix</td>
<td></td>
<td></td>
<td>(2.9, &lt;0.001)</td>
</tr>
</tbody>
</table>
Causal probabilistic network relating to a PFC module correlating with multiple LOAD clinical covariates, enriched for immune function/pathways related to microglia activity

We identified TYROBP as a key regulator of this network

CD33, MS4A4A, MS4A6A (from LOAD GWAS)

Two papers in NEJM today reporting on rare variants in TREM2 associate with LOAD
Systems analysis of eleven rodent disease models reveals an inflamatome signature and key drivers

- Atherosclerosis: ApoE KO (aorta)
- Diabetes: db/db (islet)
- Inflammatory pain: CGN (skin)
- COPD: IL-1b Tg (lung)
- Asthma: OVA model (lung)
- Fibrosis: TGFb Tg (lung)
- Diabetes: db/db (adipose)
- Obesity: ob/ob (adipose)
- Stroke: tMCAO (brain)
- Neuropathic pain: Chung (DRG)
- Inflammation: LPS (liver)
- Sarcopenia: aged rat (muscle)

<table>
<thead>
<tr>
<th>Similar set: upregulated</th>
<th>Enrichment P</th>
<th>Overlap</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory response</td>
<td>4.76E-61</td>
<td>208</td>
<td>704</td>
</tr>
<tr>
<td>Leukocyte activation</td>
<td>2.13E-32</td>
<td>164</td>
<td>704</td>
</tr>
<tr>
<td>Regulation of immune response</td>
<td>1.44E-25</td>
<td>84</td>
<td>260</td>
</tr>
<tr>
<td>Cytokine production</td>
<td>6.10E-18</td>
<td>85</td>
<td>335</td>
</tr>
<tr>
<td>Chemotaxis</td>
<td>4.97E-16</td>
<td>74</td>
<td>284</td>
</tr>
<tr>
<td>Humoral immune response</td>
<td>3.25E-14</td>
<td>69</td>
<td>271</td>
</tr>
<tr>
<td>Mitotic cell cycle</td>
<td>7.64E-13</td>
<td>87</td>
<td>414</td>
</tr>
<tr>
<td>Induction of apoptosis</td>
<td>1.74E-12</td>
<td>86</td>
<td>412</td>
</tr>
<tr>
<td>TLR signaling pathway</td>
<td>4.66E-12</td>
<td>21</td>
<td>47</td>
</tr>
<tr>
<td>Phagocytosis</td>
<td>2.74E-11</td>
<td>38</td>
<td>111</td>
</tr>
<tr>
<td>Innate immune response</td>
<td>9.29E-11</td>
<td>48</td>
<td>173</td>
</tr>
<tr>
<td>ECM remodeling</td>
<td>9.67E-11</td>
<td>22</td>
<td>59</td>
</tr>
<tr>
<td>Osteoclast differentiation</td>
<td>3.61E-10</td>
<td>31</td>
<td>82</td>
</tr>
<tr>
<td>Regulation of cell proliferation</td>
<td>4.19E-10</td>
<td>112</td>
<td>662</td>
</tr>
<tr>
<td>Antigen processing and presentation</td>
<td>7.44E-10</td>
<td>32</td>
<td>89</td>
</tr>
<tr>
<td>Positive regulation of translation</td>
<td>3.64E-09</td>
<td>45</td>
<td>170</td>
</tr>
<tr>
<td>Cytokine production by Th17 cells</td>
<td>6.33E-09</td>
<td>17</td>
<td>40</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td>9.41E-09</td>
<td>60</td>
<td>277</td>
</tr>
<tr>
<td>Cell-cycle process</td>
<td>2.79E-08</td>
<td>104</td>
<td>636</td>
</tr>
<tr>
<td>Wound healing</td>
<td>1.79E-07</td>
<td>57</td>
<td>274</td>
</tr>
<tr>
<td>Regulation of translation</td>
<td>4.44E-07</td>
<td>66</td>
<td>349</td>
</tr>
<tr>
<td>Macrophage activation</td>
<td>1.61E-06</td>
<td>20</td>
<td>49</td>
</tr>
<tr>
<td>Interleukin-12 production</td>
<td>1.70E-06</td>
<td>18</td>
<td>40</td>
</tr>
</tbody>
</table>
Core disease modules harbor pluripotent drug targets
Functional chemigenomics screen: Chemical perturbagens against disease networks *in silico*
Topiramate Reduces IBD Severity in a TNBS Rodent Model of IBD

- TNBS chemically induced rat model of IBD
- Animals treated with 80mg/kg topiramate oral after sensitization
- Prednisolone positive control (approved for IBD in humans)
Leveraging NGS and Predictive Network Models to Drive Personalized Cancer Therapy
chr17 in tumor S1 underwent somatic copy number loss

LOH of the whole chromosome 17, which includes TP53, BRCA1, CDK12, ERBB2, TRIM37

17p has an one copy loss in 77% Ovarian cancer samples in TCGA

CNV event:
0  = LOH
>0  = gain
<0  = loss

Allele imbalance:
shows which regions underwent CNV of some kind
Frameshift deletion A411fs found in CDK12 in both sites

Coverage + observed allele frequencies (if non-ref)

Read alignments showing deletion + adjacent SNV

Observed frequency of mut allele

Whole-exome seq (WES)

RNA-Seq
CDK12 primes HOW/Crn-dependent splicin in fly glial cells

Phosphorylates Ser2 in heptapeptide repeat of C-terminal domain of RNA pol II

Rodrigues F et al. Development 2012;139:1765-1776
CDK12 mutation results in loss of kinase domain

Normal: aa401 RKKKERA A AAKMDGKE SKGSPV FLPK ENS SVEAKDS...

Mutant: aa401 RKKKERA A A KQRW MERS P RVHLYFCLEKRTVQ*

NLS: nuclear localization signal
RS domain: arginine/serine-rich domain
PEST region: peptide sequence rich in proline
kinase domain: serine-threonine kinase domain
PRM: proline-rich motif

Ko TK (J Cell Sci 2001)
Chen HH (Mol Cel Biol 2006)
Tagliatala A (PhD thesis 2012)
Personalized multiscale tumor networks to diagnose and treat cancers
Key driver analysis: Identifying those genes that regulate network states that have larger impact on outcomes.

Bar chart showing fold enrichment for different categories:
- Global Drivers
- Drivers
- Local Drivers
- Module Genes
- Non-module Genes

Categories include:
- Mutation Chromatin Modification Subnetwork
- Methylation Extracellular Matrix Subnetwork
Patient mutation data projected onto the network: Interesting 1000-node subnetwork identified

Blue nodes are mutated genes

- Full network comprised of 7,881 expr/2,331 CNV nodes, 306 regulators, 501 functional mutations
- Subnetwork: 116 regulators, 232 functional mutations – massive enrichment ($p = 4.5e^{-173}$)
- 6 mutations affecting master regulators in patient and TCGA data, including ASPM and CENPF related to BUB1B dependency
- Many pathways dramatically enriched: transmembrane receptor protein tyrosine kinase activity, collagen binding, axonogenesis and so on
Using multiscale tumor networks to inform personalized chemotherapy options
Aiming to build personalized multiscale networks to model dynamics of complex disease

DNA
Cell-specific RNA
Cytokines
Clinical labs
Physiometrics
High-dimensional data acquisition carried out over time and at multiple scales can provide for the precision medicine approach we all seek.
Ultimate Objective: Predictive models to navigate your health course throughout the course of your life

Adapted from Rui Chang et al. *PLoS Computational Biology*
## Acknowledgements

<table>
<thead>
<tr>
<th>Mount Sinai</th>
<th>PacBio</th>
<th>Cornell</th>
<th>CSHL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ali Bashir</td>
<td>Jason Chin</td>
<td>Chris Mason</td>
<td>Richard McCombie</td>
</tr>
<tr>
<td>Bobby Sebra</td>
<td>Yan Gao</td>
<td>Roger Altman</td>
<td>Eric Antoniou</td>
</tr>
<tr>
<td>Joel Dudley</td>
<td>Greg Khitrov</td>
<td>Russell Durrett</td>
<td>Patricia Mocombe</td>
</tr>
<tr>
<td>Andrew Kasarskis</td>
<td>Frank Boellmann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milind Mahajan</td>
<td>Ellen Paxinos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gintaras Deikus</td>
<td>David Rank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun Zhu</td>
<td>Paul Peluso</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bin Zhang</td>
<td>Edwin Hauw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael Linderman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaurav Pandey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bojan Losic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omar Jabado</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glenn Farrell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York Genome</td>
<td>Sage Bionetworks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td></td>
<td>Stephen Friend</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chris Gaiteri</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stay connected with us! @multiscalebio